DOI:
10.1080/07391102.2018.1439403

Document Version
Peer reviewed version

Combinatorial Peptide Library Screening for Discovery of Diverse α-glucosidase Inhibitors Using Molecular Dynamics Simulations and Binary QSAR Models

Adriano Mollica,[a,*] Gokhan Zengin,[b] Serdar Durdagi,[c,d] Ramin Ekhteiari Salmas,[c] Giorgia Macedonio,[a] Azzurra Stefanucci,[a] Marilisa Pia Dimmito,[a] Ettore Novellino[e]

[a] Dipartimento di Farmacia, Università di Chieti-Pescara "G. d'Annunzio", Via dei Vestini 31, 66100 Chieti, Italy.
[b] Department of Biology, Science Faculty, Selcuk University, Konya, Turkey.
[c] Computational Biology and Molecular Simulations Laboratory, Department of Biophysics, School of Medicine, Bahcesehir University, 34349 Istanbul, Turkey.
[d]Neuroscience Program, Graduate School of Health Sciences, Bahcesehir University, Istanbul, Turkey.
[e]Dipartimento di Farmacia, Università di Napoli "Federico II", Via D. Montesano, 49, 80131 Naples, Italy.

* a.mollica@unich.it

ABSTRACT

Human α-glucosidase is an enzyme involved in the catalytic cleavage of the glucoside bond and involved in numerous functionalities of the organism, as well as in the insurgence of diabetes mellitus 2 and obesity. Thus, developing chemicals that inhibit this enzyme is a promising approach for the treatment of several pathologies. Small peptides such as di- and tri-peptides may be in natural organism as well as in the GI tract in high concentration, coming from the digestive process of meat, wheat and milk proteins. In this work, we reported the first tentative hierarchical structure-based virtual screening of peptides for human α-glucosidase. The goal of this work is to discover novel and diverse lead compounds that my act as inhibitors of α-glucosidase such as small peptides by performing a computer aided virtual screening and to find novel scaffolds for further development. Thus, in order to select novel candidates with original structure we performed molecular dynamics (MD) simulations among the 12 top-ranked peptides taking as comparison the MD simulations performed on crystallographic inhibitor acarbose. The compounds with the lower RMSD variability during the MD, were reserved for *in vitro* biological assay. The selected 4 promising structures were prepared on solid phase peptide synthesis and used for the inhibitory assay, among them compound 2 showed good inhibitory activity, which validated our method as an original strategy to discover novel peptide inhibitors. Moreover, pharmacokinetic profile predictions of these 4 peptides were also carried out with binary QSAR models using MetaCore/MetaDrug applications.

KEYWORS: α-glucosidase, inhibitors, virtual screening, drug design, peptides, binary QSAR models, MD simulations, MetaCore/MetaDrug

1. Introduction

Glucosidases are enzymes abundant in the gastrointestinal (GI) tract, where they catalyze the final step of the carbohydrates digestion. This enzyme has drawn a great interest in the scientific community since the inhibition of the glucose absorption could be used to control the blood glucose level and for hypocaloric dietary regimen. Thus, an inhibitor of α-glucosidase may be used as therapeutic agent in the treatment of diabetes and obesity. However, due to its role in the metabolism of carbohydrates it has also been speculated to be useful in the management of other diseases in which the cell metabolism has a critical role, such as cancer, viral infection, hepatitis (Rutten et al., 2006). Dietary patterns and body exercises are in the line for treatment of incipient type 2 diabetes mellitus (DM-T2) also defined as pre-diabetes. If this approach alone failed to lower the glucose concentration in the bloodstream, a medicinal approach should be recommended (Rutten et al., 2006).

Actually, six different types of antihyperglycemics are available on market: biguanides, sulphonylurea, glinidines, thiazolidinediones, dipeptidyl peptidase IV inhibitors and α-glucosidase inhibitors (Nathan, 2007). α-glucosidase inhibitors (AGIs) are reversible ligands, non specific through different maltase, and exert their effect by delaying the digestion of complex sugars in the gut, therefore, also the absorption is affected (L.K. Campbell, White and R.K. Campbell, 1996). AGIs might be an early treatment of patients with DM-T2, indeed due to their mechanism of action, acting specifically on postprandial glucose level, which is also a risk factor for cardiovascular events (Ceriello, 2005). Acarbose is the most widely prescribed AGI which is commercialized for the treatment of DM-T2, although it possesses several side effects, such as cephalea, sleeplessness, nausea, intestinal gas, and diarrhea (Van de Laar et al., 2005) (Figure 1). Among the natural products several polyphenols have α-glucosidase inhibitory effects, in fact there is a growing market of nutraceuticals for dismetabolic diseases. A great effort has been done to find novel and safe inhibitors of α-glucosidase, especially among the herbal extracts, without serious drawbacks. This area is particularly prone to be involved in the development of functional foods and nutraceuticals or to develop novel lead compounds. Several microorganisms, including e.g. *Streptomyces* sp. (Do and Joo, 1989), *Bacillus* sp. (H.S. Kim et al., 2011), *Nelumbo n.* (One, Hattori, Fukaya, Imai, Ohizumi, 2006), *Grateloupia e.* (K.Y. Kim, Nam, Kurihara, S.M. Kim, 2008), *Ganoderma l.* (S.D. Kim and H.J. Nho, 2004), and Pine bark (Y.M. Kim and H.J. Nho, 2004) have been studied to produce AGIs. Recently, some peptides derived from *Aspergillus o.* N159-1 (Kang, Yi, Soo, 2013) have been studied as inhibitors of α-glucosidase, indeed several peptides able to modulate the activity of α-glucosidase, have been found in a series of proteins from silkworm and phage peptides (Watanabe et al., 2013). Although these peptides are generally less potent than synthetic drugs they have a strong potentiality as active components for food fortification and as diet integrators. In this paper, we have identified a novel class of α-glucosidase peptides inhibitors by performing a computer aided drug-design protocol. Our experiments involved the building of a general purpose virtual peptide library containing around 10000 peptides (after preparation). In a further step this peptide library was screened by docking simulations, the hits were rescored by MM/GBSA and molecular dynamics (MD) simulations. The selected peptides were synthesized and tested on α-glucosidase and the obtained results have been used to validate our approach.

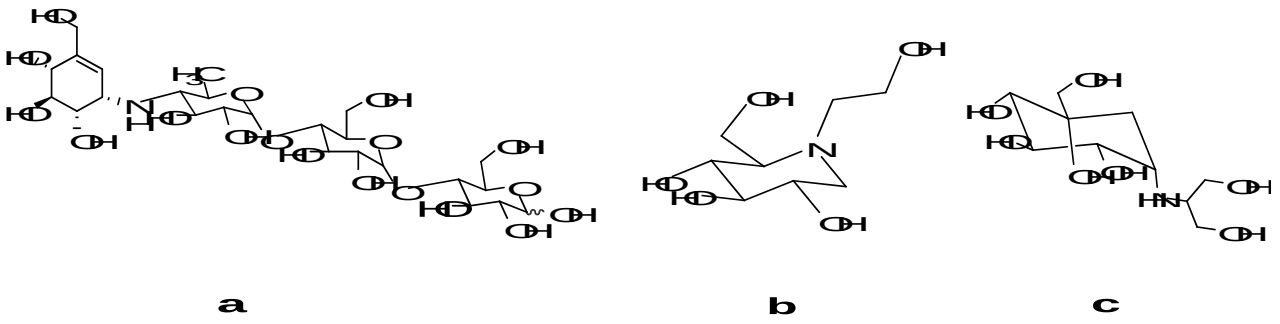


Figure 1. AGI inhibitors: **a)** Acarbose; **b)** Miglitol; **c)** Voglibose.

2. Materials and Methods

2.1. *Peptide Library Generation*

The peptide library has been built *in house*, by a python program which generates all possible combinations of strings containing three letters among the 20 one-letter code of the amino acids. By this way, the program generated 20^3 novel combinations of tripeptides without protonation states, which became around 10000 after ligand preparation being included also alternative protonation states

for some residues, e.g. for Histidines. These obtained strings database was converted in 3D chemical structures using Maestro molecular modeling suit (Schrödinger Release 2017-4: Maestro, Schrödinger, LLC, New York, NY, 2017). Then the peptide library needs to be prepared and manually amidated at the *C*-terminus.

2.2. *Protein Preparation*

The crystal structure of α-glucosidase (PDB ID, 3TOP) was obtained from the RCSB database (Shen, Qin, Ren, 2011; Joint Center for Structural Genomics, Crystal structure of Alpha-glucosidase (TM0752) from Thermotoga maritima at 2.50 A resolution RCSBPSP, (2004) 1-14). This structure needs to be prepared by protein preparation wizard (Sastry, Adzhigirey, Day, Annabhimoju and Sherman, 2013) tool embedded in Maestro molecular modeling suite. The PropKa program was used to locate protonation states at the physiological pH of 7.4 (Olsson, Søndergaard, Rostkowski and Jensen, 2011). Hydrogen atoms were added and minimized. In the raw pdb file, other molecules also present in the crystallization medium, these molecules have been removed.

2.3. *Glide Docking of the co-crystallized ligand docking*

In an early stage, we tested the suitability of Glide extra precision (XP) protocol to predict the ligand position of the reference compound (Friesner et al., 2006). Co-crystallized inhibitor acarbose was prepared by the LigPrep module, and then it was docked to the active sites of α-glucosidase by Glide/XP protocol. The binding cavity was assigned as an outer box of size 25x 25 x25 Å, determined by the position of the crystallographic ligand. The RMSD value for acarbose was calculated by SID (simulation interface) and compared crystallized acarbose after XP docking and was found as less than 1.3 Å (Figure 2).

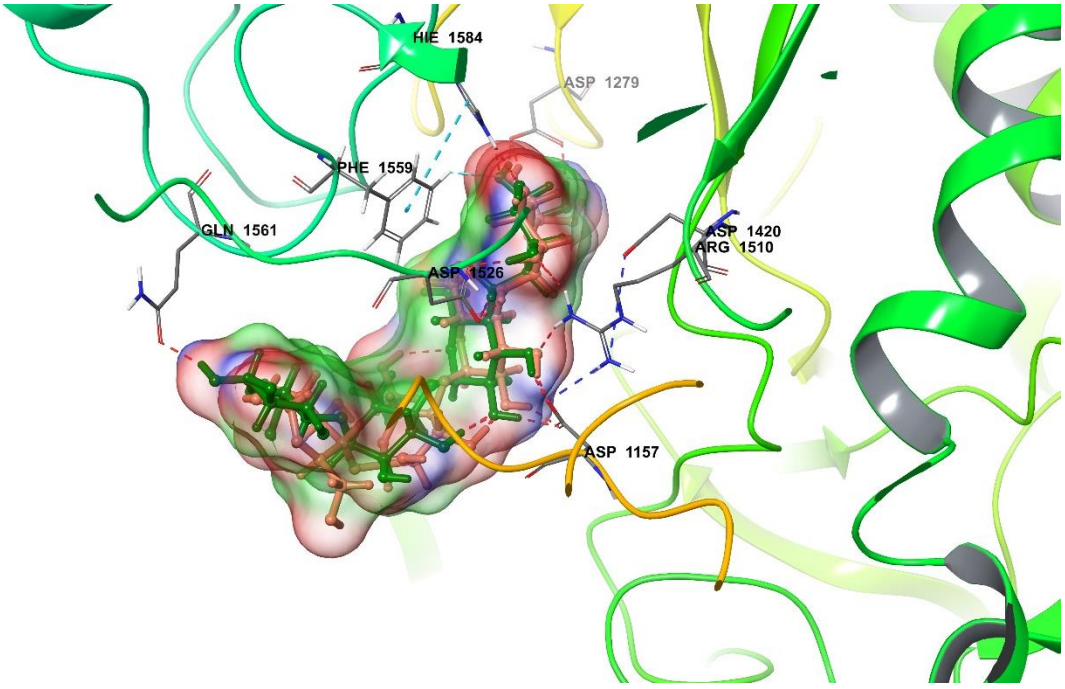


Figure 2. Acarbose crystallographic ligand superimposed to Acarbose pose obtained after Glide docking.

2.4. *Peptide Library Docking*

Docking experiments were performed by Glide (Friesner et al., 2006). The grid to set up the docking was created with standard parameters optimized for peptide docking. The grid box coordinates were

centered at the crystallographic ligand and a size of 25.0 Å of each side was assigned. The peptide library was prepared by LigPrep tool (Sastry et al., 2013), the molecules ionized at a pH of 7.4 by Epik. The number of peptide ligands, including tautomers rise up to around 10000. The library was used after minimization by OPLS-3 ff without further modifications.

2.5. *Virtual screening workflow*

Docking with XP scoring method within Glide has been adopted to dock the library into the acarbose defined enzymatic pocket with the default parameters, and the best ranked 3300 (30% of the molecules giving a scoring value of at least -8.0 kcal/mol) ranking poses of each molecule were retained for the next step of hits enrichment, by Prime/MM-GSBA (Bell et al., 2012). Finally, 20 ns molecular dynamic (MD) simulations were performed by using the docking poses of that compounds with the best free binding energy. The MD simulation trajectory analysis led us to identify the final hits for in vitro analysis. MD simulations allow to study the connections of the best ranked pose found by Glide and ranked by Prime, as well as to observe the time-dependent evolution of the interactions between ligand and enzyme, not detectable from the docking study only.

2.6. *Prime/MM-GBSA*

The top 30% of the poses obtained by docking od the entire library with Glide/XP scoring function were retained and submitted for MM/GBSA calculations performed with Prime module of Maestro. All the parameters were kept default. MM/GBSA is able to calculate the free binding energy of the complex ligand-enzyme in a water environment. The energy was estimated by the OPLS3 force field (Harder et al., 2016) for molecular mechanic energy (E_{MM}) and the surface-generalized Born model (SGBM) for polar solvation energy (VSGB), and the apolar solvation factor (G_{SA}). This energetic term is calculated for the complex, for the ligand and for the protein alone, then the binding free energy (ΔG_{bind}) is finally obtained by the following equation:

$$G = E_{MM} + G_{SGB} + G_{SA}$$

$$\Delta G_{bind} = G_{complex} - (G_{protein} - G_{ligand})$$

The top-12 compounds ranked by the free binding energy returned by MM/GBSA experiments were submitted for MD (Figure 3).

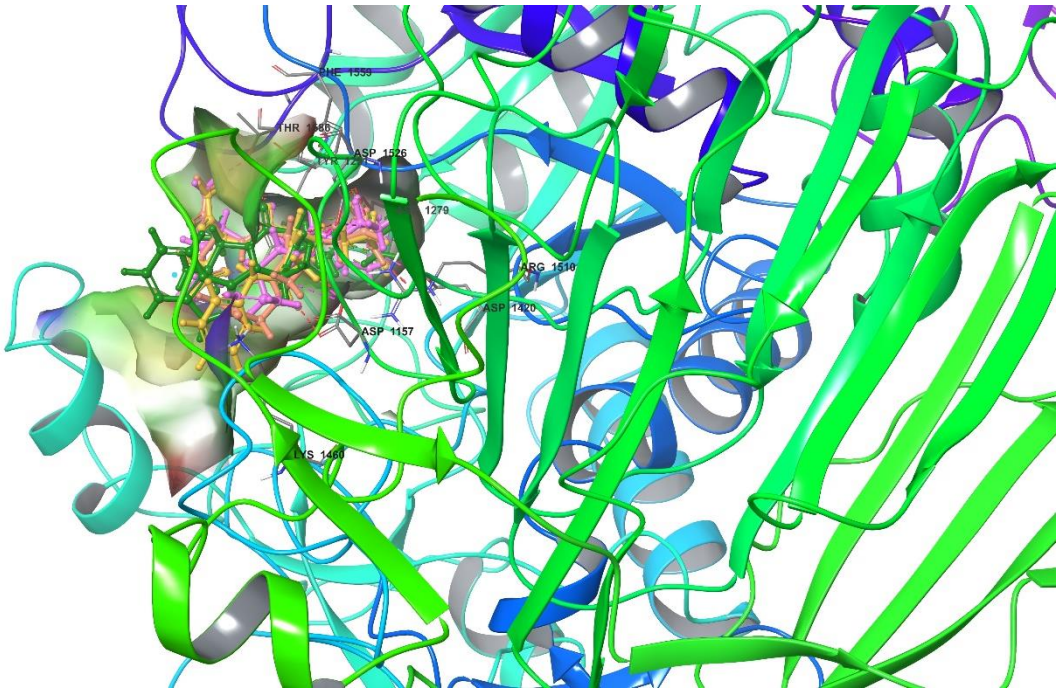


Figure 3. Superimposing of the four ligands (**1-4**) selected after MD simulations.

2.7. *Molecular Dynamics (MD) simulations*

The docked tripeptide-enzyme poses were subjected to MD simulations in water (Hospital, Goñi, Orozco and Gelpì, 2015). Desmond has been employed for the simulations. The complex was inserted in aqueous environment. The model constitutes of a box of water in which is inserted the complex. The box has a minimum size as to contain the complex ensuring a distance of 10 Å from the edge of the box and the protein. In order to neutralize the system, 0.15 M NaCl has been added to the system. The OPLS3 ff was used for all the experiments. TIP3P-TIP4P model (Penna and Lennart, 2001) was used water. The system was minimized up to 2000 steps, holding all the protein and ligand atoms. The minimized system was subjected to MD simulations, using the NPT ensemble and periodic boundary conditions for 20 ns. To control the pressure Martyna-Tobias-Klein method was used which allows to keep the pressure of the system at 1.01 bar by using the isotropic coupling method. The Nose-Hoover thermostat was applied to control the temperature at 310K. The trajectories and other parameters were saved every 20 and 1.2 ps respectively to returns 1000 frames. The simulation analysis was done by the simulation interactive diagram (SID), especially useful to visualize the RMSD fluctuations of ligand and enzyme, hydrogen bonds stability, rotation of groups on the ligand, van der Waals interactions over the simulation trajectories and overall stability of the secondary structure of the enzyme.

2.8. *MetaCore/MetaDrug Applications*

1. The hit molecules obtained from docking and MD simulations protocol were evaluated for the investigation of their therapeutic activity, pharmacokinetic and toxicity properties by using MetaCore/MetaDrug comprehensive systems biology analysis suite of Clarivate Analytics with the help of available ADME, disease and toxicity QSAR models (Mollica, Costante, Akdemirb et al., 2015). MetaCore is based on a high-quality, manually-curated database of molecular interactions, molecular pathways, gene-disease associations, chemical metabolism and toxicity information. The hit molecules obtained from virtual screening were tested for their biological effects as potential AGI candidates. Furthermore, various toxic effects such as mutagenecity, anemia, carcinogenicity, cardiotoxicity, cytotoxicity, epididymis toxicity, genotoxicity, hepatotoxicity, kidney necrosis, kidney growth, liver

cholestasis, etc. were predicted by 26 different toxicity QSAR models under MetaCore (https://portal.genego.com/).

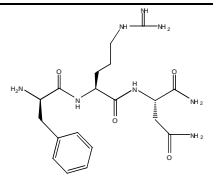
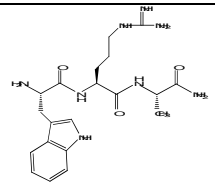
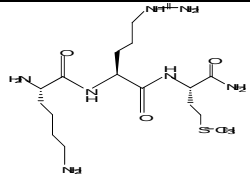
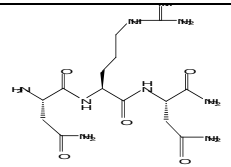
COMPOUNDS	SEQUENCE	STRUCTURES	Docking score (kcal/mol)	ΔG binding free energy (kcal/mol)	MS calculated	MS found
1	H-Phe-Arg-Asn-NH_2		-9.363	-69.896	434.24	435.1
2	H-Trp-Arg-Ala-NH_2		-8.108	-72.861	430.24	431.3
3	H-Lys-Arg-Met-NH_2		-9.569	-69.328	432.26	433.5
4	H-Asn-Arg-Asn-NH_2		-9.094	-79.613	401.21	402.4

Table 1. Structures of compounds **1-4**, docking scores, ΔG free energy, MS data.

3. Chemistry

3.1. Synthesis of compounds

The synthesis of compounds **1–4** was performed by solid phase peptide synthesis (SPPS)(Mollica, Pinnen, Stefanucci and Costante, 2013; Mollica, Costante, Novellino et al., 2015) (Table 1). The N terminus protecting group was Fmoc, the protection group of the side chains of the amino acids were tert-butyloxy-carbonyl for Lysine and Tryptophan, Pbf for Arginine, Trityl for Asn. We used Rink amide resin with a 1.2 mMol/g loading, to obtain amidates peptides at the C terminus, the system TBTU/HOBt/DIPEA was used as coupling mixture and piperidine 20% in DMF for Fmoc group deprotection as previously reported by our group (Scheme 1). Purification via RP-HPLC afforded the desired peptides in excellent yields. Peptides were characterized by ^{1}H NMR and Ms spectrometer.

3.2. Materials and Methods

Solvents and reagents were purchased from WVR (Italy) and used as supplied without further purification. Amino acids were purchased from GLS Shanghai (China). ^{1}H NMR spectra were recorded at 25° C on a 300 MHz Varian Mercury spectrometer. Chemical shifts are reported in parts per million (δ) downfield from the internal standard TMS. Mass spectra were recorded on a LCQ Finnigan-Mat mass spectrometer (San Jose, CA) by ESI-spay source and ion trap analyzer. The capillary temperature was set at 200 °C and the spray voltage at 4.00 kV. The electropray was formed by using nitrogen (N_2) as both the sheath gas and the auxiliary gas with a flow of elium. The purity of final TFA salts was confirmed by NMR analysis, ESI-LRMS, and analytical RP-HPLC recorded at 216, 235 254, and 275 nm (Waters C18 4.6 mm × 150 mm) at a flow rate of 1 mL/min, using as eluent a gradient of H_2O/acetonitrile−0.1% TFA ranging from 5% ACN to 90% ACN in 32 minutes, and was found to be≥96%.

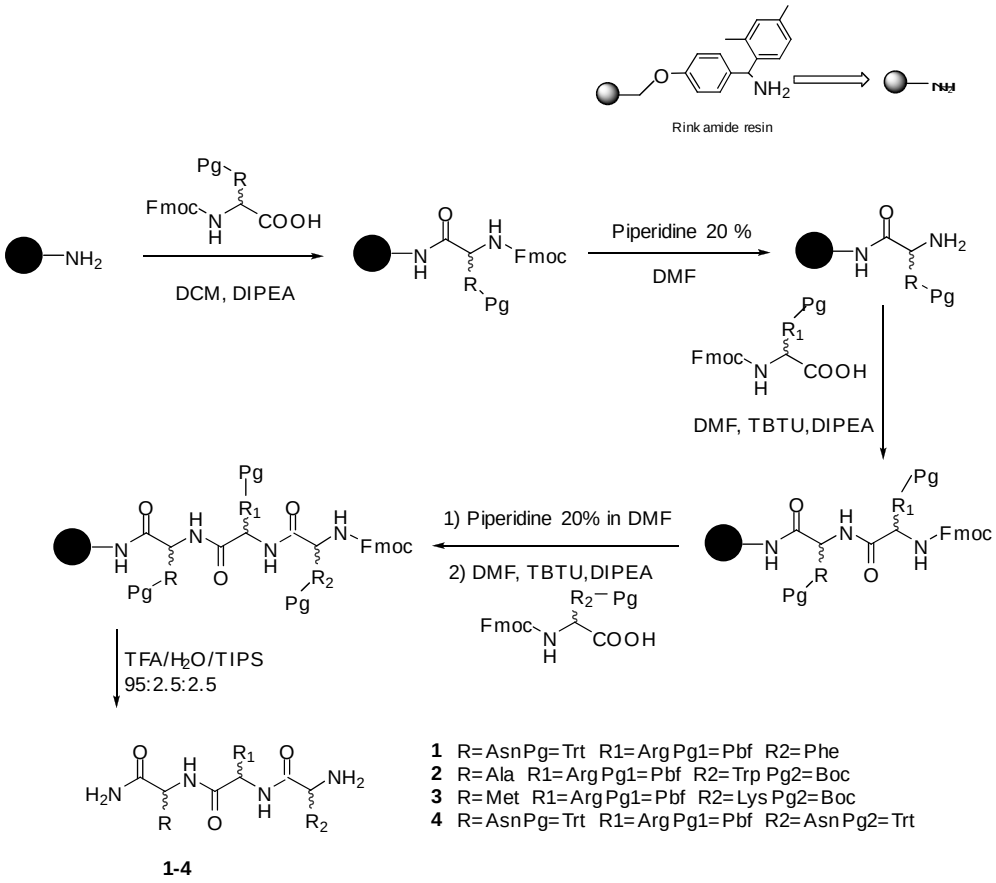
3.3. Synthesis

3.3.1. Resin loading

The Rink Amid Fmoc resin was deprotected by 20% piperidine in DMF (2 × 15 min). The resin was flush washed by treatment with MeOH/DMF/DCM repeated three times each. The first amino acid to load (3 equiv) is dissolved in a flask in DMF (5 mL), then TBTU (3 equiv) and DIPEA (6 equiv) were added and the resulting mixture was shaked for 10 minutes, then added to the resin. The coupling was kept shacked vigorously overnight. Kaiser test was routinely used to check the coupling. After competition, the resin was flush washed with MeOH/DMF/DCM three times each.

3.3.2. Cleavage and purification

In order to cleave the peptide from the resin, a mixture of TFA/H_2O/TIPS 95:2.5:2.5 (8 mL for 1 h) was added to the plastic vessel and shacked for 2.5 hours. After this time the solution is collected, concentrated under vacuum and the resulting ˜1 mL of solution, that was precipitated in cold ether. The peptide was separated from the solution by centrifugation at 4000 rpm for 5 minutes, and the ether removed. This process war repeated four times, the obtained solid was then dried at reduced pressure. Final peptides **1-4** were purified by HPLC using a Waters XBridge semiprep C18, 5.0 μm, 250 mm× 10 mm column at a flow rate of 4 mL/min on a Waters pump 600, using as eluent a linear gradient of H_2O+0.1% TFA/ACN+0.1% TFA, from 5% ACN to 90% ACN in 32 min.



Scheme 1. Preparation of products **1-4** via SPPS.

4. Enzyme Assay

4.1. a-glucosidase inhibition

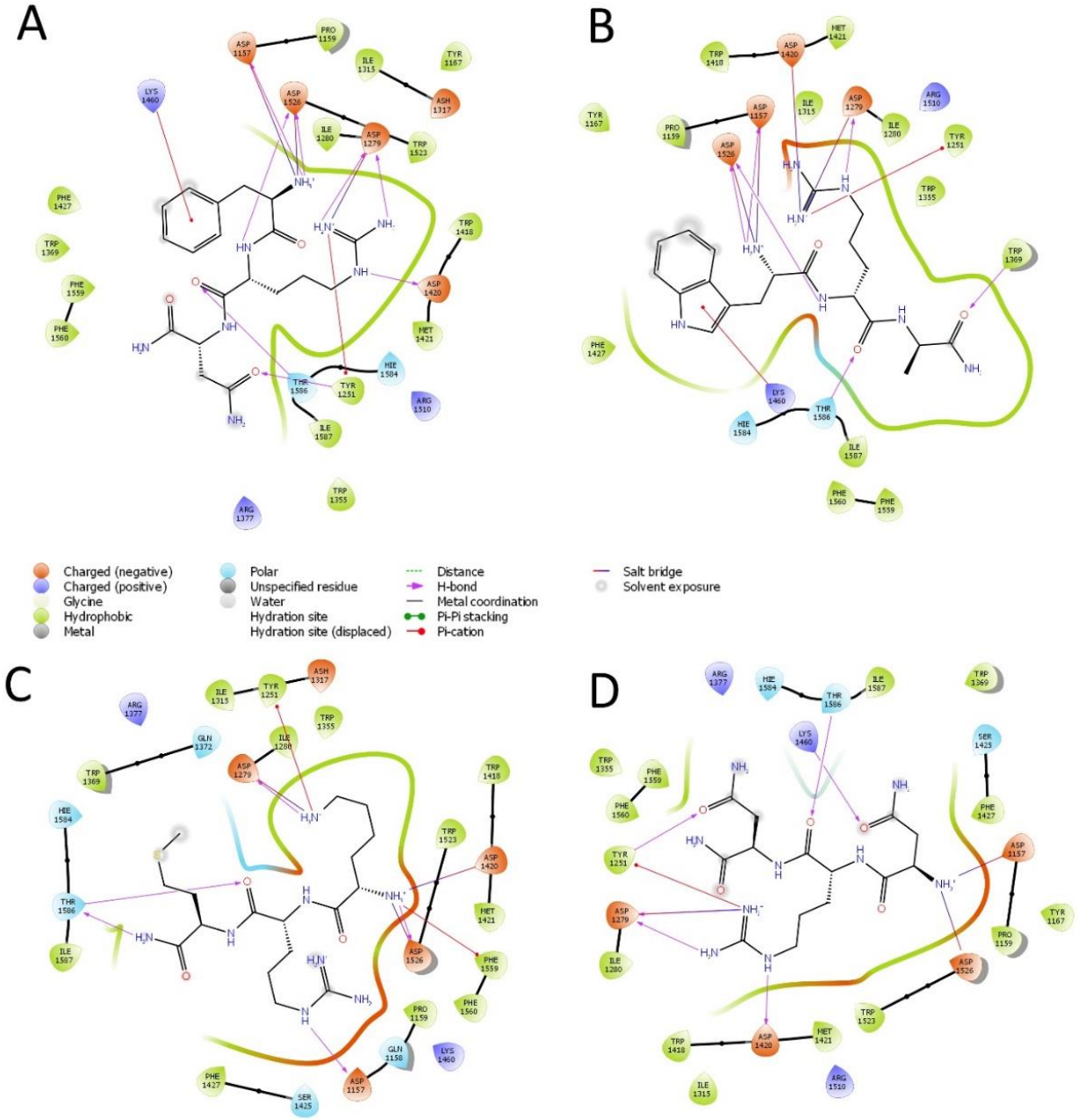
The four peptides **1-4** were tested for inhibition of α-glucosidase as previously reported by our research group (Zengin, 2016) (Table 2). Compound solution (50 µL) was mixed with GSH (50 µL), α-glucosidase stock (from Saccharomyces cerevisiae, EC 3.2.1.20, Sigma) (50 µL) in buffer (phosphate, pH 6.8) and PNPG (4-N-trophenyl-α-D-glucopyranoside) (50 µL) in a 96-well microplate and incubated for 15 min at 37 °C. Similarly, a blank was prepared by adding compound solution to all reaction reagents without enzyme (α-glucosidase) solution. The reaction was then stopped with the addition of Na_2CO_3 (50 µL, 0.2 M). The sample and blank absorbances were read at 400 nm. The absorbance of the blank was subtracted from that of the sample inhibitory activity of each compounds on α-glucosidase activity was calculated as equivalents of acarbose.

Peptides	Metal chelating (mgEDTAE/g sample)	Glucosidase Inhibition (µmolACAE/g sample)
Compound 1	5,63±0.20	0,45±0.08
Compound 2	5,58±0.43	0,70±0.09
Compound 3	5,46±1.29	0,58±0.05
Compound 4	3,86±0.18	0,36±0.08

Table 2. Biological assay of compounds **1-4**.

5. Results and discussion

VS on glucosidase has been proved to be a valuable tool to find new lead compounds (Sanders et al., 2012). To discover novel inhibitor candidates of α-glucosidase, a virtual peptide library was generated *in house* and screened *in silico*. The virtual screening was carried out into three steps, firstly all the ligands were screened by docking performed by Glide/XP scoring function (Halgren et al., 2004). In a second step, around 30% of the ligands with high-docking score have been retained for MM/GBSA re-scoring. The top-10 hits were retained for MD simulations. After RMSD analysis, four peptides have been selected for synthesis by SPPS methodology and tested by *in vitro* analysis in order to validate the virtual screening results. *In vitro* analysis showed that compounds **1-4** inhibited α-glucosidase with an inhibitory potency calculated in equivalents of acarbose reported in Table 2. Compound **2** was found to be the most active among this series. Several papers deal with the developing of suitable computational methods of scoring accurately the molecules contained in virtual libraries generally composed of 'small molecules' (Sliwoski, Kothiwale, Meiler and Lowe, 2014); however, at the best of our knowledge this is the first attempt to use a combinatorial peptides' library for a virtual screening. We used Glide/XP to produce the docking poses, as it was found to be able to reproduce the pose of the crystallographic ligand acarbose, and the best 30% compounds were used for rescoring by MM/GBSA, which in several papers is reported to be one of the most reliable method of rescoring the hits after docking. Prime/MM-GBSA has been applied taking in consideration its good performance to rank the potential ligands better than Glide. This quality has been validated in several targets, which is probably due to its realistic parametrization of the solvation that occurs to the complex in water. The satisfactory combination of Glide to generate reliable binding poses joined to accuracy of Prime in the calculation of the free binding energy produced 12 best hits, that were reduced to four after MD simulations. As illustrated in Figure 4, the four inhibitor candidates have similar predicted docking poses in the binding to α-glucosidase (Figure 1A), and their highly flexible structures, fit well into the acarbose-binding pocket, due to a similarity in steric and electronic distribution. The potency of the novel AGIs was in micro-molar range, however none of the compounds reached the docking score and MM/GBSA calculated for acarbose, in agreement with the activity observed for the four peptides (the most active peptide (**2**) has an activity equal to 0.7 equivalents of acarbose).


A
B
C
D
Charged (negative)
Charged (positive)
Glycine
Hydrophobic
Metal
Polar
Unspecified residue
Water
Hydration site
Hydration site (displaced)
Distance
H-bond
Metal coordination
Pi-Pi stacking
Pi-cation
Salt bridge
Solvent exposure

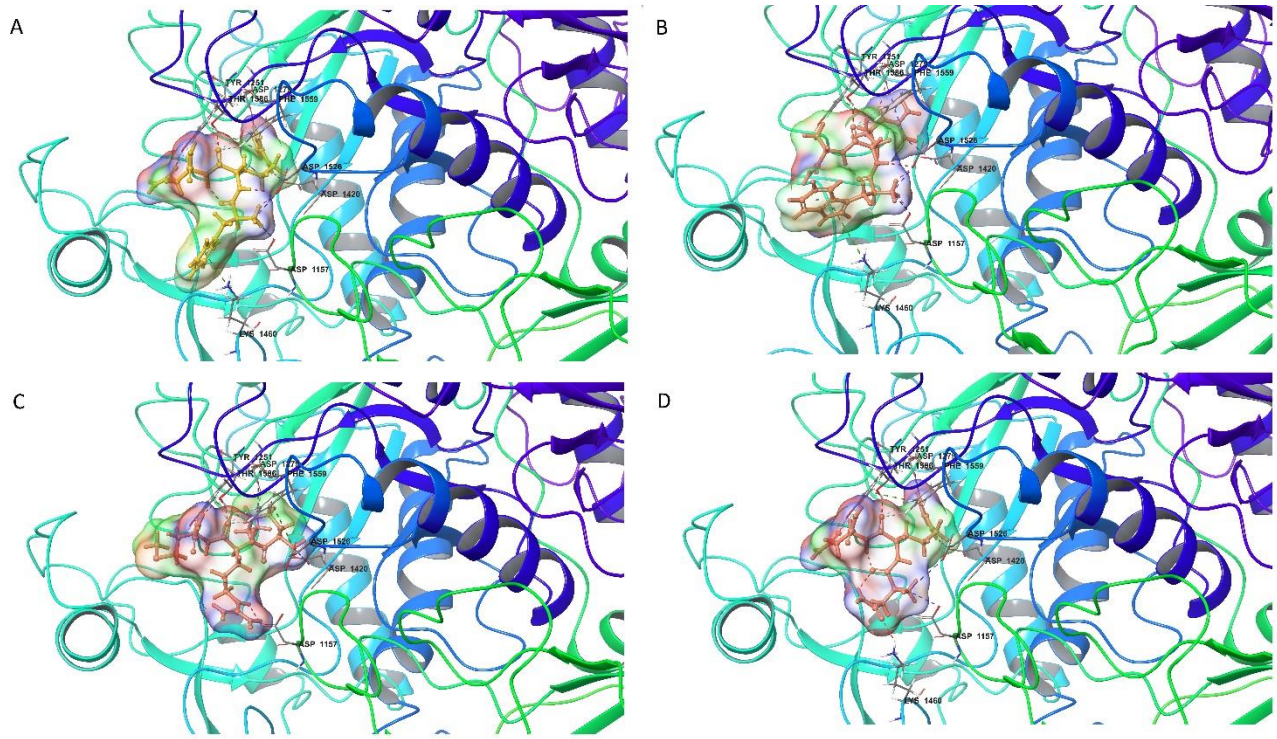


Figure 4. Docking poses of compounds **1-4** into the binding pocket of α-glucosidase.

The crystallographic ligand acarbose interacts with α-glucosidase in the enzymatic cavity by establishing several hydrogen-bond (H-bond) interactions. In particular, the key interactions involve: Asp1157, Asp1279, Lys1460, Arg1510, Asp1526, His1584. The peptide inhibitors contain several hydrophilic groups, such as the amide and amine groups, which may mimicry the polar interactions with the polar region of the binding site in a similar way of acarbose. However, not all H-bond interactions made by acarbose are respected, such as for the H-bond toward Arg1510, His1584 that were not observed for compounds **2** and **3**. All the tested peptides (compounds **1-4**) share the following interactions: Asp1157, Asp1279, Asp1526. Peptide **4** has also in common with acarbose the H-bond to Lys1460, whereas the compounds **1** and **2** do not display this H bond but interact to the same residue by establishing a π-cation interaction. Peptide **3** does not interact with Lys1460 residue but forms a π-cation interaction with Phe1559. Moreover, all the selected compounds have further interaction to Tyr1251 (π -cation), Asp1420 (ionic for **2** and **3** and H-bond for **1** and **4**), Thr1586 (H-bond). Peptide **2** which has shown the best activity, may provide a starting lead structure for further optimizations as selective inhibitor. The observed preferential inhibition of the compound **2** also provides helpful information about the preferential binding mode and key interactions. The best pose found for **2** indeed showed the presence of a folded motif, namely a γ-turn inverse turn (Guruprasad and Rajkumar, 2000; Mollica et al., 2006; Aschi et al., 2003; Morera et al., 2005; Durdagi et al., 2013), centered at the lysine residues and stabilized by the intramolecular H-bond formed with the interaction of the carbonyl oxygen of the Trp residue (i) and the proton of the amide group of the alanine residue (i+2) Figure 5.

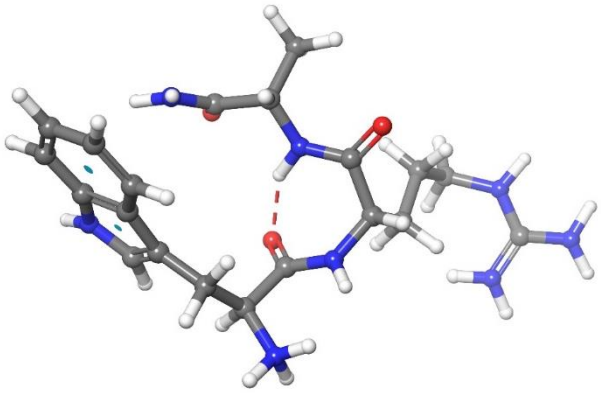


Figure 5. Peptide **2**, intramolecular H bond stabilizes a ß-turn structure.

MetaCore/MetaDrug Applications

When Tables 3-5 are examined it can be seen that selected peptides can have some therapeutic potentials (cut-off value for therapeutic potential is 0.5). The values in brackets in this table is *Tanimato Prioritization (TP)*. A maximal Tanimato coefficient calculated for all molecules in a training set of a QSAR model indicates similarity of the analyzed structure to the most similar compound in the training set. The high docking scores and the acceptability of the values of therapeutic activities alone may not be enough for the molecules we choose. Taking this into account we have investigated the ADME properties like BBB (blood brain barrier), G-log P (lipophilicity), Prot-bind, % (connectivity to blood serum proteins) and Prot-bind, log t (connectivity to blood serum albumin proteins). These parameters are very important for the candidate drug-like molecules (Table 3). ADME/QSAR models, therapeutic and toxicity prediction profiles of 4 peptides with positive control molecules are compared. Based on results, lipophilicity (log of compound octanol/water distribution) values of designed peptides are found high (Table 4, Table 5). Although blood brain barrier penetration of 4 peptides are smaller than cutoff value (-0.3), their values are similar/smaller than positive controls. Within the 4 peptides, peptides 1, 2 and 3 showed potential antidiabetic activities (Table 4).

ADME QSAR Models

	Peptide-1	Peptide-2	Peptide-3	Peptide-4	Acarbose	Voglibose	Miglitol
BBB, log ratio [1]	-0.88 (38.92)	-0.81 (59.20)	-1.02 (45.16)	-0.84 (43.10)	-0.93 (44.70)	-1.00 (32.69)	-1.24 (36.91)
G-LogP [2]	0.79	1.55	1.04	0.33	0.68	-1.01	-1.44
Prot-bind, % [3]	40.96 (58.50)	51.38 (47.34)	37.02 (45.16)	37.02 (45.45)	52.66 (58.33)	21.48 (41.33)	20.23 (44.44)
Prot-bind, Log t [4]	-0.13 (38.92)	0.05 (59.20)	-0.14 (45.16)	-0.33 (43.10)	-0.47 (44.70)	-0.43 (32.69)	-0.38 (36.91)
WSol, log mg/L [5]	1.65	1.97	2.27	2.28	3.31	4.17	5.08

Table 3. ADME QSAR Models.

1. Blood brain barrier penetration model. The data is expressed as log values of the ratio of the metabolite concentrations in brain and plasma. Cutoff is -0.3. Larger values indicate that the metabolite is more likely to enter the brain. Model description: N=107, R2=0.89, RMSE=0.26.

2. Lipophilicity, log of compound octanol-water distribution. Cutoffs are -0.4 to 5.6. Values greater than 5.6 correspond to overly hydrophobic compounds. Model description: N=13474, R2=0.95, RMSE=0.21.

3. Human serum protein binding, %. Cutoff is 50%. A value of more than 95% is highly bound, less than 50% is a low binding metabolite. Model description: N=265, R2=0.909, RMSE=10.11.

4. Affinity to human serum albumin, log value of the retention time. Cutoff is 0. Positive values correspond to higher protein binding, negative values to lower protein binding. An acceptable level of binding is project dependent. The model is based on retention times of compounds assayed by HPLC using an immobilized HSA column. Values are expressed as log values of the retention time. Model description: N=95, R2=0.904, RMSE=0.2.

5. Water solubility at 25°C, log mg/L. Cutoffs are from 2 to 4. An acceptable level of solubility is project dependent. Model description: N=2871, 2=0.91, RMSE=0.54.

Prediction of Therapeutic Activity

	Peptide-1	Peptide-2	Peptide-3	Peptide-4	Acarbose	Voglibose	Miglitol
Allergy [1]	0.17 (65.54)	0.17 (51.80)	0.33 (43.16)	0.33 (54.46)	0.30 (52.23)	0.42 (42.86)	0.44 (38.13)
Alzheimer [2]	0.58 (71.23)	0.45 (79.21)	0.47 (48.82)	0.65 (58.82)	0.43 (42.98)	0.67 (48.19)	0.54 (50.00)
Angina [3]	0.28 (54.50)	0.32 (69.86)	0.32 (57.72)	0.30 (69.81)	0.08 (60.54)	0.29 (46.15)	0.47 (41.18)
Arthritis [4]	0.85 (60.10)	0.96 (76.09)	0.82 (44.72)	0.71 (50.00)	0.74 (34.25)	0.46 (46.15)	0.33 (42.11)
Asthma [5]	0.64 (67.35)	0.60 (76.09)	0.51 (57.72)	0.45 (69.81)	0.50 (48.11)	0.64 (48.19)	0.66 (50.00)
Bacterial [6]	0.22 67.35)	0.24 (67.51)	0.72 (50.53)	0.74 (59.26)	0.67 (63.01)	0.81 (44.80)	0.84 (45.90)
Cancer [7]	0.76 (55.28)	0.76 (67.59)	0.86 (57.72)	0.79 (69.81)	0.17 (100.00)	0.84 (71.62)	0.90 (90.77)
Depression [8]	0.45 (66.42)	0.54 (81.01)	0.52 (47.20)	0.47 (57.41)	0.07 (51.93)	0.16 (61.90)	0.16 (65.00)
Diabetes [9]	0.56 (71.64)	0.42 (79.35)	0.57 (42.62)	0.55 (48.15)	0.34 (100.00)	0.63(100.00)	0.46 (78.26)
HIV[10]	0.81 (68.03)	0.74 (76.09)	0.69 (47.20)	0.65 (57.41)	0.46 (58.33)	0.76 (48.19)	0.67 (50.00)
Heart Failure [11]	0.95 (66.92)	0.90 (82.90)	0.84 (45.16)	0.84 (45.76)	0.65 (50.21)	0.34 (41.27)	0.36 (40.98)
Hyperlipidemia [12]	0.46 (51.72)	0.55 (42.01)	0.47 (37.67)	0.48 (37.50)	0.58 (55.75)	0.38 (43.48)	0.34 (43.06)
Hypertension [13]	0.76 (63.41)	0.53 (65.05)	0.54 (49.01)	0.63 (59.29)	0.36 (60.54)	0.53 (77.14)	0.68 (98.36)
Inflammation [14]	0.36 (51.72)	0.22 (61.83)	0.41 (50.39)	0.47 (57.89)	0.16 (100.00)	0.37 (44.76)	0.20 (50.51)
Migraine [15]	0.66 (68.03)	0.60 (76.09)	0.73 (57.72)	0.65 (69.81)	0.52 (58.33)	0.69 (51.25)	0.63 (55.26)
Mycosis [16]	0.38 (71.64)	0.26 (65.05)	0.34 (44.68)	0.39 (44.44)	0.30 (56.28)	0.39 (40.94)	0.49 (36.18)
Obesity [17]	0.91 (60.10)	0.97 (76.09)	0.68 (41.73)	0.55 (50.00)	0.85 (48.11)	0.28 (46.02)	0.05 (42.11)
Osteoporosis [18]	0.88 (60.10)	0.87 (76.09)	0.27 (49.33)	0.28 (51.00)	0.48 (48.11)	0.47 (46.02)	0.64 (42.11)
Pain [19]	0.37 (56.69)	0.35 (82.90)	0.37 (50.85)	0.45 (62.38)	0.04 (60.54)	0.50 (42.70)	0.67 (41.35)
Parkinson [20]	0.53 (67.35)	0.43 (81.01)	0.38 (38.51)	0.47 (43.54)	0.30 (52.23)	0.60 (38.46)	0.48 (35.00)
Psoriasis [21]	0.42 (65.54)	0.29 (67.51)	0.28 (49.01)	0.29 (45.76)	0.33 (46.64)	0.26 (39.20)	0.27 (35.71)
Schizophrenia [22]	0.44 (48.05)	0.38 (65.05)	0.30 (52.99)	0.29 (63.37)	0.35 (56.04)	0.32 (39.20)	0.30 (36.89)
Skin Diseases [23]	0.49 (56.04)	0.35 (47.09)	0.54 (45.30)	0.59 (39.19)	0.35 (56.04)	0.75 (36.94)	0.58 (43.06)
Thrombosis [24]	0.65 (56.25)	0.47 (65.05)	0.47 (53.15)	0.58 (60.82)	0.29 (58.33)	0.24 (40.86)	0.19 (38.89)
Viral [25]	0.48 (71.64)	0.68 (56.57)	0.42 (48.82)	0.51 (47.24)	0.28 (50.23)	0.53 (40.86)	0.57 (48.44)

Table 4. Prediction of therapeutic activities.

1. Potential antiallergic activity. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs. Model description: Training set N=258, Test set N=47, Sensitivity= 0.87, Specificity=0.88, Accuracy=0.87, MCC=0.74.

2. Potential activity against Alzheimer's disease. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs, drug candidates in clinical trials and preclinical compounds with in vivo activity. Model description: Training set N=261, Test set N=44, Sensitivity= 0.91, Specificity=0.82, Accuracy=0.86, MCC=0.73.

3. Potential antianginal activity. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs, drug candidates in clinical trials and preclinical compounds with in vivo activity. Model description: Training set N=546, Test set N=95, Sensitivity= 0.90, Specificity=0.93, Accuracy=0.92, MCC=0.83.

4. Potential activity against arthritis. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs, drug candidates in clinical trials and preclinical compounds with in vivo activity. Model description: Training set N=460, Test set N=77, Sensitivity= 0.98, Specificity=0.94, Accuracy=0.96, MCC=0.92.

5. Potential activity against asthma. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs, drug candidates in clinical trials and preclinical compounds with in vivo activity. Model description: Training set N=366, Test set N=63, Sensitivity= 0.92, Specificity=0.86, Accuracy=0.89, MCC=0.78.

6. Potential antibacterial activity. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs. Model description: Training set N=530, Test set N=97, Sensitivity= 0.87, Specificity=0.90, Accuracy=0.89, MCC=0.77.

7. Potential activity against cancer. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs. Model description: Training set N=886, Test set N=167, Sensitivity= 0.89, Specificity=0.83, Accuracy=0.86, MCC=0.72.

8. Potential activity against depression. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs. Model description: Training set N=335, Test set N=62, Sensitivity= 0.93, Specificity=0.82, Accuracy=0.87, MCC=0.75.

9. Potential antidiabetic activity. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs, drug candidates in clinical trials and preclinical compounds with in vivo activity. Model description: Training set N=195, Test set N=34, Sensitivity= 0.85, Specificity=0.93, Accuracy=0.88, MCC=0.77.

10. Potential activity against HIV. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs, drug candidates in clinical trials and preclinical compounds with in vivo activity. Model description: Training set N=491, Test set N=80, Sensitivity= 0.80, Specificity=0.86, Accuracy=0.84, MCC=0.67.

11. Potential activity against heart failure. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs. Model description: Training set N=204, Test set N=33, Sensitivity= 0.78, Specificity=0.87, Accuracy=0.82, MCC=0.64.

12. Potential antihyperlipidemic activity. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs. Model description: Training set N=185, Test set N=24, Sensitivity= 0.75, Specificity=0.92, Accuracy=0.83, MCC=0.68.

13. Potential antihypertensive activity. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs. Model description: Training set N=554, Test set N=111, Sensitivity= 0.89, Specificity=0.81, Accuracy=0.85, MCC=0.70.

14. Potential anti-inflammatory activity. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs. Model description: Training set N=598, Test set N=93, Sensitivity= 0.86, Specificity=0.84, Accuracy=0.85, MCC=0.69.

15. Potential activity against migraine. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs, drug candidates in clinical trials and preclinical compounds with in vivo activity. Model description: Training set N=515, Test set N=98, Sensitivity= 0.81, Specificity=0.84, Accuracy=0.83, MCC=0.65.

16. Potential antifungal activity. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs. Model description: Training set N=172, Test set N=47, Sensitivity= 0.90, Specificity=0.88, Accuracy=0.89, MCC=0.79.

17. Potential activity against obesity. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs, drug candidates in clinical trials and preclinical compounds with in vivo activity. Model description: Training set N=472, Test set N=75, Sensitivity= 0.89, Specificity=0.97, Accuracy=0.93, MCC=0.87.

18. Potential anti-osteoporosis activity. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs, drug candidates in clinical trials and preclinical compounds with in vivo activity. Model description: Training set N=595, Test set N=86, Sensitivity= 0.84, Specificity=0.85, Accuracy=0.85, MCC=0.70.
19. Potential analgetic activity. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs. Model description: Training set N=525, Test set N=84, Sensitivity= 0.92, Specificity=0.67, Accuracy=0.79, MCC=0.60.
20. Potential activity against Parkinson's disease. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs, drug candidates in clinical trials and preclinical compounds with in vivo activity. Model description: Training set N=298, Test set N=49, Sensitivity= 0.96, Specificity=0.96, Accuracy=0.96, MCC=0.92.
21. Potential activity against psoriasis. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs, drug candidates in clinical trials and preclinical compounds with in vivo activity. Model description: Training set N=199, Test set N=32, Sensitivity= 0.93, Specificity=0.82, Accuracy=0.89, MCC=0.74.
22. Potential activity against schizophrenia. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs, drug candidates in clinical trials and preclinical compounds with in vivo activity. Model description: Training set N=616, Test set N=93, Sensitivity= 0.89, Specificity=0.91, Accuracy=0.90, MCC=0.80.
23. Potential activity against skin diseases. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs. Model description: Training set N=255, Test set N=36, Sensitivity= 1.00, Specificity=0.76, Accuracy=0.86, MCC=0.76.
24. Potential antithrombotic activity. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs, drug candidates in clinical trials and preclinical compounds with in vivo activity. Model description: Training set N=453, Test set N=80, Sensitivity= 0.98, Specificity=0.95, Accuracy=0.97, MCC=0.93.
25. Potential antiviral activity. Cutoff is 0.5. Values higher than 0.5 indicate potentially active compounds. Training set consists of approved drugs. Model description: Training set N=206, Test set N=35, Sensitivity= 0.92, Specificity=0.95, Accuracy=0.94, MCC=0.88.

Prediction of Toxic Effects

	Peptide-1	Peptide-2	Peptide-3	Peptide-4	Acarbose	Voglibose	Miglitole
AMES [1]	0.64 (54.14)	0.50 (43.10)	0.52 (41.18)	0.50 (47.73)	0.55 (50.62)	0.43 (41.27)	0.32 (44.26)
Anemia [2]	0.35 (49.65)	0.28 (46.08)	0.32 (45.16)	0.36 (45.45)	0.09 (56.73)	0.56 (39.20)	0.42 (44.44)
Carcinogenicity [3]	0.16 (71.64)	0.08 (82.90)	0.37 (69.61)	0.46 (82.76)	0.16 (54.47)	0.90 (60.56)	0.95 (63.24)
Carcinogenicity Mouse Female [4]	0.12 (71.64)	0.11 (82.90)	0.31 (46.92)	0.30 (63.89)	0.20 (54.47)	0.72 (60.56)	0.75 (63.24)
Carcinogenicity Mouse Male [5]	0.28 (71.64)	0.15 (82.90)	0.40 (69.61)	0.31 (82.76)	0.17 (54.47)	0.69 (60.56)	0.56 (63.24)
Carcinogenicity Rat Female [6]	0.04 (71.64)	0.03 (82.90)	0.17 (50.00)	0.25 (63.89)	0.09 (54.47)	0.79 (43.33)	0.81 (42.67)
Carcinogenicity Rat Male [7]	0.06 (71.64)	0.06 (82.90)	0.35 (50.00)	0.37 (58.95)	0.17 (54.47)	0.89 (41.74)	0.86 (40.51)
Cardiotoxicity [8]	0.20 (39.16)	0.11 (46.08)	0.15 (40.14)	0.25 (45.45)	0.86 (48.42)	0.81 (39.20)	0.90 (36.91)
Cytotoxicity Model, -log GI50 (M) [9]	4.82 (52.98)	5.12 (58.75)	4.42 (48.36)	4.78 (59.05)	5.01 (55.94)	5.28 (48.82)	4.55 (46.46)
Epididymis Toxicity [10]	0.20 (51.70)	0.24 (65.05)	0.34 (41.12)	0.34 (50.55)	0.38 (44.88)	0.62 (37.80)	0.67 (43.06)
Genotoxicity [11]	0.33 (65.54)	0.37 (82.90)	0.46 (49.33)	0.54 (47.73)	0.31 (54.47)	0.92 (41.33)	0.75 (38.41)
Hepatotoxicity [12]	0.17 (69.62)	0.14 (92.49)	0.26 (53.23)	0.28 (60.00)	0.59 (54.68)	0.56 (100.00)	0.48 (78.26)
Kidney Necrosis [13]	0.15 (47.14)	0.06 (55.36)	0.34 (40.14)	0.55 (45.45)	0.17 (55.75)	0.87 (41.74)	0.63 (38.84)
Kidney Weight Gain [14]	0.10 (60.56)	0.09 (92.49)	0.36 (50.00)	0.32 (58.95)	0.42 (58.33)	0.85 (69.57)	0.52 (62.86)
Liver Cholestasis [15]	0.24 (58.50)	0.15 (55.36)	0.15 (45.16)	0.29 (45.45)	0.43 (49.05)	0.63 (100.00)	0.66 (78.26)
Liver Lipid Accumulation [16]	0.28 (60.56)	0.24 (92.49)	0.26 (40.31)	0.29 (47.24)	0.37 (48.80)	0.46 (48.19)	0.60 (42.86)
Liver Necrosis [17]	0.49 (40.91)	0.21 (54.65)	0.50 (41.78)	0.38 (38.06)	0.32 (50.62)	0.83 (41.75)	0.81 (41.58)
Liver Weight Gain [18]	0.41 (48.17)	0.05 (59.26)	0.54 (36.50)	0.47 (37.30)	0.40 (54.10)	0.95 (100.00)	0.74 (78.26)
MRTD [19]	0.37 (51.55)	0.26 (58.22)	0.69 (53.23)	0.64 (60.00)	0.98 (59.57)	1.14 (78.26)	1.31 (100.00)
Nasal Pathology [20]	0.07 (51.18)	0.09 (48.88)	0.21 (39.31)	0.11 (46.88)	0.20 (35.08)	0.39 (34.21)	0.70 (35.62)
Nephron Injury [21]	0.29 (53.85)	0.02 (55.22)	0.32 (49.33)	0.47 (47.24)	0.19 (58.33)	0.77 (48.19)	0.77 (44.44)
Nephrotoxicity [22]	0.10 (53.85)	0.13 (55.22)	0.35 (49.33)	0.40 (59.81)	0.36 (58.33)	0.57 (100.00)	0.66 (100.00)
Neurotoxicity [23]	0.26 (48.28)	0.12 (67.86)	0.41 (36.36)	0.35 (38.78)	0.25 (47.78)	0.86 (44.25)	0.86 (44.44)
Pulmonary Toxicity [24]	0.06 (71.64)	0.05 (67.59)	0.11 (43.69)	0.12 (54.02)	0.22 (61.71)	0.29 (51.25)	0.56 (55.26)
SkinSens, EC3 [25]	28.63 (31.62)	43.94 (51.06)	27.88 (37.89)	13.69 (44.44)	49.35 (24.11)	4.19 (33.33)	4.73 (28.00)
Testicular Toxicity [26]	0.46 (51.70)	0.59 (65.05)	0.43 (45.45)	0.36 (50.55)	0.22 (54.10)	0.42 (52.38)	0.49 (55.00)

Table 5. Prediction of toxicity values.

1. Potential to be mutagenic (AMES positive), range from 0 to 1. A value of 1 is AMES positive (mutagenic), and a value of 0 is AMES negative (non-mutagenic). Cutoff is 0.5. Values close to zero are preferable. The AMES assay is based upon the reversion of mutations in the histidine operon in the bacterium Salmonella enterica sv Typhimurium.
2. Potential for causing anemia. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Training set consists of chemicals and drugs causing anemia in vivo. Model organisms: human. Model description: Training set N=324, Test set N=51, Sensitivity= 0.82, Specificity=0.90, Accuracy=0.86, MCC=0.72.

3. Potential for inducing carcinogenicity in rats and mice. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Training set consists of chemicals and drugs causing carcinogenicity in vivo. Model organisms: mouse, rat. Model description: Training set N=1210, Test set N=185, Sensitivity= 0.96, Specificity=0.90, Accuracy=0.93, MCC=0.86.
4. Potential for inducing carcinogenicity in female mice. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Training set consists of chemicals and drugs causing carcinogenicity in vivo. Model organisms: female mice. Model description: Training set N=640, Test set N=94, Sensitivity= 0.90, Specificity=0.93, Accuracy=0.92, MCC=0.83.
5. Potential for inducing carcinogenicity in male mice. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Training set consists of chemicals and drugs causing carcinogenicity in vivo. Model organisms: mouse male. Model description: Training set N=584, Test set N=93, Sensitivity= 0.91, Specificity=0.88, Accuracy=0.89, MCC=0.78.
6. Potential for inducing carcinogenicity in female rats. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Training set consists of chemicals and drugs causing carcinogenicity in vivo. Model organisms: female rat. Model description: Training set N=667, Test set N=120, Sensitivity= 0.90, Specificity=0.96, Accuracy=0.93, MCC=0.86.
7. Potential for inducing carcinogenicity in male rats. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Training set consists of chemicals and drugs causing carcinogenicity in vivo. Model organisms: male rat. Model description: Training set N=715, Test set N=117, Sensitivity= 0.92, Specificity=0.88, Accuracy=0.90, MCC=0.79.
8. Potential for inducing cardiotoxicity. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Training set consists of chemicals and drugs causing cardiotoxicity in vivo. Model organisms: mouse, rat, human. Model description: Training set N=143, Test set N=30, Sensitivity= 0.80, Specificity=1.00, Accuracy=0.90, MCC=0.82.
9. Growth inhibition of MCF7 cell line (human caucasian breast adenocarcinoma), pGI50. Cutoff is 6. Values from 6 to 8 correspond to a toxic metabolite, values less than 6 are preferable, values less than 3 are more preferable and less toxic. Model description: N=1474, R2=0.9, RMSE=0.05.
10. Potential for inducing epididymis toxicity. Training set consists of chemicals and drugs causing epididymis toxicity in vivo. Model organisms: mouse, rat, human. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Model description: Training set N=252, Test set N=42, Sensitivity= 0.90, Specificity=0.86, Accuracy=0.88, MCC=0.76.
11. Potential for inducing genotoxicity. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Training set consists of chemicals and drugs causing genotoxicity in vivo. Model organisms: mouse, rat. Model description: Training set N=372, Test set N=86, Sensitivity= 0.75, Specificity=0.84, Accuracy=0.79, MCC=0.59.
12. Potential for inducing hepatotoxicity. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Training set consists of chemicals and drugs causing hepatotoxicity in vivo. Model organisms: mouse, rat, human. Model description: Training set N=1380, Test set N=231, Sensitivity= 0.73, Specificity=0.88, Accuracy=0.81, MCC=0.62.
13. Potential for inducing kidney necrosis. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Training set consists of chemicals and drugs causing renal necrosis in vivo. Model organisms: mouse, rat, human. Model description: Training set N=221, Test set N=42, Sensitivity= 0.96, Specificity=1.00, Accuracy=0.98, MCC=0.95.
14. Potential for inducing kidney weight gain. Cutoff is 0.5. The values higher than 0.5 indicate potentially toxic compounds. Training set consists of chemicals and drugs causing kidney weight gain in vivo. Model organisms: mouse, rat. Model description: Training set N=240, Test set N=49, Sensitivity= 0.95, Specificity=1.00, Accuracy=0.98, MCC=0.96.
15. Potential for inducing liver cholestasis. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Training set consists of chemicals and drugs causing cholestasis in vivo. Model organisms: mouse, rat, human. Model description: Training set N=218, Test set N=35, Sensitivity= 0.79, Specificity=0.67, Accuracy=0.74, MCC=0.46.
16. Potential for inducing liver lipid accumulation. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Training set consists of chemicals and drugs causing lipid accumulation in vivo. Model organisms: mouse, rat, human. Model description: Training set N=172, Test set N=28, Sensitivity= 0.80, Specificity=0.85, Accuracy=0.82, MCC=0.64.
17. Potential for inducing liver necrosis. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Training set consists of chemicals and drugs causing hepatic necrosis in vivo. Model organisms: mouse, rat, human. Model description: Training set N=300, Test set N=57, Sensitivity= 0.91, Specificity=0.91, Accuracy=0.91, MCC=0.82.
18. Potential for inducing liver weight gain. Cutoff is 0.5. Values higher than 0.5 indicate potential liver weight-changing compounds. Training set consists of chemicals and drugs causing liver weight gain in vivo. Model organisms: mouse, rat. Model description: Training set N=292, Test set N=52, Sensitivity= 1.00, Specificity=1.00, Accuracy=1.00, MCC=1.00.
19. Maximum Recommended Therapeutic Dose, log mg/kg-bm/day, range is from -5 to 3. Cutoff is 0.5. Chemicals with high log MRTDs can be classified as mildly toxic compounds, chemicals with low log MRTDs as highly toxic compounds. Model description: N=1209, R2= 0.86, RMSE=0.42.
20. Potential for causing nasal pathology. Training set consists of chemicals and drugs causing nasal pathology in vivo. Model organisms: mouse, rat, human. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Model description: Training set N=246, Test set N=47, Sensitivity= 1.00, Specificity=0.93, Accuracy=0.96, MCC=0.92.
21. Potential for inducing nephron injury. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Training set consists of chemicals and drugs causing nephron injury in vivo. Model organisms: mouse, rat, human. Model description: Training set N=598, Test set N=109, Sensitivity= 0.91, Specificity=1.00, Accuracy=0.96, MCC=0.93.
22. Potential for inducing nephrotoxicity. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Training set consists of chemicals and drugs causing nephrotoxicity in vivo. Model organisms: mouse, rat, human. Model description: Training set N=847, Test set N=154, Sensitivity= 0.90, Specificity=0.84, Accuracy=0.87, MCC=0.74.
23. Potential for inducing neurotoxicity. Training set consists of chemicals and drugs causing neurotoxicity in vivo. Model organisms: mouse, rat, human. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Model description: Training set N=175, Test set N=34, Sensitivity= 0.94, Specificity=0.94, Accuracy=0.94, MCC=0.88.
24. Potential for inducing pulmonary toxicity. Training set consists of chemicals and drugs causing pulmonary toxicity in vivo. Model organisms: mouse, rat, human. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Model description: Training set N=482, Test set N=87, Sensitivity= 0.89, Specificity=0.88, Accuracy=0.89, MCC=0.77.
25. Skin sensitization potential expressed as effective concentration 3, EC3 %. Values higher than 10 indicate weak and moderate sensitizers. Model description: N=89, R2=0.67, RMSE=22.56.
26. t consists of chemicals and drugs causing testicular toxicity in vivo. Model organisms: mouse, rat, human. Cutoff is 0.5. Values higher than 0.5 indicate potentially toxic compounds. Model description: Training set N=439, Test set N=88, Sensitivity= 0.81, Specificity=0.85, Accuracy=0.83, MCC=0.66.

6. CONCLUSION

In summary, four potential α-glucosidase tripeptides inhibitors have been found by applying a hierarchic approach of virtual screening (Aschi et al., 2003). The novel peptides discovered, if compared with acarbose, possess a completely novel scaffold and showed distinct interaction modes in the acarbose-catalytic cavity to some extent. Results of binary QSAR models showed that synthesized peptides have similar/better pharmacokinetic and phramacodynamic properties compared to tested three known α-glucosidase inhibitors. Their importance is of great interest since the α-glucosidase

activity could be certainly mudulated by the plethora of peptide fragments present in the GI tract during the digestion of proteins. The distinct conformations of the most active peptide could help to enhance our knowledge, furnishing further models useful for the design of new α-glucosidase inhibitors. Also, taking in consideration that small peptides are naturally occurring substances, present in the GI tract in huge amount during digestion of native proteins, we consider that the leads found during this study may open the way to a new paradigm in the study of α-glucosidase modulation, including dietary peptides and proteins, may concur to develop new tools to fight against type 2 diabetes.

CONFLICT OF INTEREST

Declared none.

REFERENCES

Aschi, M., Lucente, G., Mazza, F., Mollica, A., Morera, E., Nalli, M., Paglialunga Paradisi, M. (2003). Peptide backbone folding induced by the C(alpha)-tetrasubstituted cyclic alpha-amino acids 4-amino-1,2-dithiolane-4-carboxylic acid (Adt) and 1-aminocyclopentane-1-carboxylic acid (Ac5c). A joint computational and experimental study. *Org. Biomol. Chem.* 1, 1980-1988.

Bell, J. A., Cao, Y., Gunn, J. R., Day, T., Gallicchio, E., Zhou, Z., ... Farid, R. (2012). PrimeX and the Schrödinger Computational Chemistry Suite of Programs. *International Tables for Crystallography*, 18, 534-538.

Campbell, L. K., White, J. R. and Campbell, R. K. (1996). Acarbose: its role in the treatment of diabetes mellitus. *Ann. Pharmacother.* 30, 1255-1262.

Ceriello, A. (2005). Postprandial hyperglycemia and diabetes complications: is it time to treat?. *Diabetes,* 54, 1-7.

Do, J. H. and Joo, H. K. (1989) α-D-Glucosidase inhibitor from Streptomyces sp. (I) Identification of the strain. *Korean J Appl Microbiol Bioeng.* 17, 202-206.

Durdagi, S., Scozzafava, G., Vullo, D., Sahin, H., Kolayli, S. and Supuran, C. T. (2013). Inhibition of mammalian carbonic anhydrases I-XIV with grayanotoxin III: solution and in silico studies. *Journal of Enzyme Inhibition and Medicinal Chemistry* 29, 469-475.

Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., ... Mainz, D. T. (2006). Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.*, 49, 6177-6196.

Guruprasad, K. and Rajkumar, S. (2000). Beta-and gamma-turns in proteins revisited: a new set of amino acid turn-type dependent positional preferences and potentials. *J. Biosci.* 25, 143-156.

Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T. and Banks, J. L. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.*, 47, 1750-1759.

Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, Y. J., ... Friesner, R. A. (2016). OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.*, 12, 281-296.

Hospital, A., Goñi, J. R., Orozco, M. and Gelpí, J. L. (2015). Molecular dynamics simulations: advances and applications. *Appl. Bioinform. Chem.*, 8, 37-47.

Joint Center for Structural Genomics, Crystal structure of Alpha-glucosidase (TM0752) from Thermotoga maritima at 2.50 A resolution RCSBPSP, (2004) 1-14.

Kang, M. G. Yi, S. H. and Soo, J. (2013). Production and Characterization of a New α-Glucosidase Inhibitory Peptide from Aspergillus oryzae N159-12013. *Mycobiology*, 41, 149-154.

Kim, H. S., Lee, J. Y., Hwang, K. Y., Cho, Y. S., Park, Y. S., Kang, K. D. and Seong, S. I. (2011). Isolation and Identification of Bacillus sp. producing α-glucosidase Inhibitor 1-deoxynojirimycin. *Korean J. Microbiol. Biotechnol.* 39, 49–55.

Kim, K. Y., Nam, K. A., Kurihara, H. and Kim, S. M. (2008). Potent α-glucosidase inhibitors purified from the red alga Grateloupia elliptica. *Phytochemistry* 69, 2820-2825.

Kim, Y. M., Wang, M. H. and Rhee, H. I. (2004). A novel α-glucosidase inhibitor from pine bark. *Carbohydr Res.* 339, 715-717.

Kim, S. D. and Nho, H. J. (2004). Isolation and characterization of α-glucosidase inhibitor from the fungus Ganoderma lucidum. *J. Microbiol.* 42, 223-227.

Mollica, A., Costante, R., Akdemirb, A., Carradori, S., Stefanucci, A., Macedonio, G., ... Supuran, C. T. (2015) Exploring new Probenecid-based carbonic anhydrase inhibitors: Synthesis, biological evaluation and docking studies. *Bioorganic & Medicinal Chemistry*, 23, 5311-5318.

Mollica, A., Costante, R., Novellino, E., Stefanucci, A., Pieretti, S., Zador, F., … Lewis, R.J. (2015). Design, Synthesis and Biological Evaluation of Two Opioid Agonist and Cav2.2 Blocker Multitarget Ligands. *Chemical Biology and Drug Design 86,* 156-162.

Mollica, A., Paglialunga Paradisi, M., Torino, D., Spisani, S. and Lucente, G. (2006). Hybrid alpha/beta-peptides: for-Met-Leu-Phe-OMe analogues containing geminally disubstituted beta2,2- and beta 3,3-amino acids at the central position. *Amino Acids* 30, 453-459.

Mollica, A., Pinnen, F., Stefanucci, A. and Costante, R. (2013). The Evolution of Peptide Synthesis: From Early Days to Small Molecular Machines. *Current Bioactive Compounds* 9, 184-202.

Morera, E., Nalli, M., Mollica, A., Paglialunga Paradisi, M., Aschi, M., Gavuzzo, E., ... Lucente, G. (2005). Peptides containing 4-amino-1,2-dithiolane-4-carboxylic acid (Adt): conformation of Boc-Adt-Adt-NHMe and NH...S interactions. *J. Pept. Sci.* 11, 104-112.

Nathan, D. M. (2007). Finding new treatments for diabetes - how many, how fast... how good?. *N. Engl. J. Med.* 356, 437-440.

Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. and Jensen, J. H. (2011). PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem Theory Comput.*, 7, 525-537.

Ono, Y., Hattori, E., Fukaya, Y., Imai, S. and Ohizumi, Y. (2006). Anti-obesity effect of Nelumbo nucifera leaves extract in mice and rats. *J. Ethnopharmacol.* 106, 238-244.

Pekka, M. and Lennart, N. (2001). Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. M. *J. Phys. Chem.* 105, 9954-9960.

Rutten, G. E. H. M, De Grauw, W. J. C., Nijpels, Houweling, B., van de Laar, F., Henk Bilo, G., ... Jansenn, P. (2006). Dutch College of General Practitioners: guidelines on Type 2 diabetes, second revision (in Dutch), Huisarts Wet. 49, 137-152.

Sanders, M. P. A., Roumen, L., van der Horst, E., Lane, J. R., Vischer, H. F., van Offenbeek, J., ... de Graaf, C. (2012). A Prospective Cross-Screening Study on G-Protein-Coupled Receptors: Lessons Learned in Virtual Compound Library Design. *J. Med. Chem.*, 55 5311-5325.

Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R. and Sherman, W. (2013) Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aid. Mol. Des.*, 27, 221-234.

Schrödinger Release 2017-4: Maestro, Schrödinger, LLC, New York, NY, 2017.

Shen, Y., Qin, X. H. and Ren L. M. (2011). Crystral Structure of the C-terminal Subunit of Human Maltase-Glucoamylase in Complex with Acarbose. *RCSBPBP*, 1-46.

Sliwoski, G., Kothiwale, S., Meiler, J. and Lowe, E. W. (2014). Computational Methods in Drug Discovery. *Pharmacol Rev.* 66, 334-395.

Van de Laar, F. A., Lucassen, P. L., Akkermans, R. P., Van de Lisdonk, E. H., Rutten, G. E. and Van Weel, C. (2005). Alpha-glucosidase inhibitors for patients with type 2 diabetes: results from a Cochrane systematic review and meta-analysis, Diabetes Care, 28 (2005) 154–163.

Watanabe, S., Kakudoc, K., Ohtab, M., Mitad, K., Fujiyamac, K. and Inumarub, S. (2013). Molecular cloning and characterization of the α-glucosidase II from Bombyx mori and Spodoptera frugiperda, Insect. *Biochemistry and Molecular Biology* 43, 319-327.

Zengin, G. (2016). A study on in vitro enzyme inhibitory properties of asphodeline anatolica: new sources of natural inhibitors for public health problems. *Industrial Crops and Products* 83, 39-43.